Defense Horizons
Number 30
A publication of the
Center for Technology and National Security Policy
National Defense University
JULY 2003

Moore's Law: A Department of Defense Perspective

by Gerald M. Borsuk and Timothy Coffey

Overview

The past 50 years have seen enormous advances in electronics and the systems that depend upon or exploit them. The Department of Defense (DOD) has been an important driver in, and a profound beneficiary of, these advances, which have come so regularly that many observers expect them to continue indefinitely. However, as Jean de la Fontaine said, "In all matters one must consider the end." A substantial literature debates the ultimate limits to progress in solid-state electronics as they apply to the current paradigm for silicon integrated circuit (IC) technology. The outcome of this debate will have a profound societal impact because of the key role that silicon ICs play in computing, information, and sensor technologies.

The consequences for DOD are profound. For example, DOD planning assumptions regarding total situational awareness have been keyed to Moore's Law, which predicts the doubling of transistor density about every 18 months. While this prediction proved to be accurate for more than thirty years, we are entering a period when industry will have increasing difficulty in sustaining this pace. Under the current device and manufacturing paradigm, progress in areas such as total situational awareness will slow or stagnate. If DOD planning assumptions are to be met, the DOD science and technology program would be well advised to search aggressively for alternate paradigms beyond those on which Moore's Law is based to ensure new technology capabilities. The purpose of this paper is to examine the current prognosis for silicon IC technology from a DOD perspective.

The Current Situation

The integrated circuit electronics revolution can be said to have begun on February 23, 1940, when Russell Ohl of Bell Laboratories observed anomalous behavior of the electronic properties of a cracked silicon crystal. His investigation led to the discovery of what is now known as the *pn junction*. Ohl's interest was in developing a better crystal oscillator. He has commented that Bell Laboratories managers were not especially interested in his work and preferred that he focus on issues related to vacuum electronics, where the real opportunities were perceived to lie. Fortunately, Walter Brattain was one of the first to review Ohl's discovery. Consequently, Bell Laboratories undertook a program to produce a solid-state switch to replace vacuum tube amplifiers and unreliable mechanical relays necessary for telephony. This program led to the discovery of the transistor in December 1947 by Brattain, John Bardeen, and William Shockley. In 1958, Jack Kilby invented and demonstrated an elemental integrated circuit composed of resistors and an active transistor device. Robert Noyce independently invented another form of the integrated circuit based upon silicon planar technology. At that point, the stage was set for the scientific and technical revolution in solid-state electronics that produced the tremendous capabilities in electronics, computers, communications, and information technology that we are experiencing today.

In 1965, Gordon Moore predicted that the number of active transistor devices on a silicon integrated circuit would double about every 12 months.[1] He based this prediction upon a log-linear plot of device complexity over time using just three empirical data points from his employer, Fairchild Semiconductor Corporation. In 1975, Moore revisited this topic at the Institute of Elecrical and Electronics Engineers International Electron Devices Meeting. At that time, (presumably with knowledge of the technical attributes of silicon metal-oxide semiconductor [MOS] device scaling[2] and his own observations of improvements in silicon planar manufacturing technologies, including economy of scale and batch processing of silicon wafer), Moore revised his prediction, stating that transistor density would double about every 18 months. This prediction became known as Moore's Law. It must be remembered that this "law" is actually an empirical prediction, not a law of nature.

The semiconductor industry established Moore's Law as a goal in the development of ICs. The information technology industry uses it as a predictive tool to allow efficient planning of investments. Thus, Moore's Law became a self-fulfilling prophecy for the past 30 years.

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
JUL 2003	**N/A**	**-**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Moore's Law: A Department of Defense Perspective	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **National Defense University Center for Technology and National Security Policy Fort McNair Washington, DC 20319**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release, distribution unlimited

13. SUPPLEMENTARY NOTES
The original document contains color images.

14. ABSTRACT

15. SUBJECT TERMS

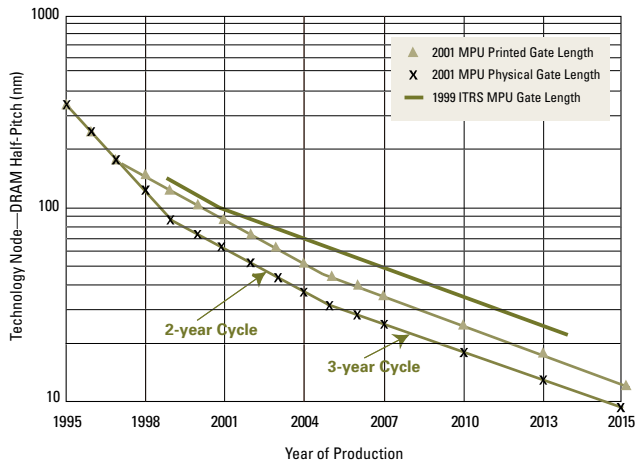
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
unclassified	**unclassified**	**unclassified**	**UU**	**8**	

Figure 1 plots integrated circuit gate feature size as a function of time in recent years and projects it into the future. The number of transistors that can be placed upon a given unit area increases as the inverse square of the feature size due to the technical attributes of device scaling. The ability of the semiconductor industry to reduce feature size continuously underlies the validity of Moore's Law and the enormous advances that have occurred in semiconductor electronics over the past 40 years. These advances in silicon integrated circuit electronics form the foundation of the great advances in computer capability and information and sensor technologies. To the extent that feature size reduction (and the resolution of all associated issues, such as other scaling attributes, on-chip interconnects, and manufacturing technologies) continues, Moore's Law will accurately predict these trends. If feature size reduction slows or fails for any reason, Moore's Law fails as a predictor, unless a viable paradigm other than feature size reduction is found. The question becomes that of determining the prognosis for continued feature size reduction and, if there are problems in this regard, assessing the prognosis for alternative paradigms. The stakes are high because we have progressed only about halfway on a log plot through the available feature size space of an integrated circuit, leaving available another three orders of magnitude in physical size reduction (assuming that the size of an atom sets the ultimate limit). Clearly, some signs of future difficulties are apparent. Three definitive regions can be discerned in the curves of figure 1. Through 1999, Moore's Law, on an 18-month cycle, prevailed. The slope starting in 1999 and predicted by the International Technology Roadmap for Semiconductors (ITRS) roadmap to continue to 2005 is one of a 2-year cycle time. From 2005 to 2016, a 3-year cycle time is projected.

DOD has depended on rapid advances in electronics of all types (for example, digital, radio frequency, mixed signal, and electro-optic) to maintain technological superiority. The department is expecting advances in electronics to continue this technological superiority indefinitely. However, potential problems are beginning to appear, specifically with digital ultra large silicon integrated (ULSI) circuits. To understand this, one must examine the paradigm on which Moore's Law is based. The law is rooted in the attributes of device scaling most simply expressed as the reduction in transistor active gate feature size and the so-called batch semiconductor manufacturing technologies necessary to make these ULSI circuits economically. A high yield of circuits with ever-increasing function complexity and performance has been the desired goal.

Device scaling. Silicon has played a predominant role in the progress of integrated circuits. This element is found in abundance in the Earth's crust, is relatively easy to purify and form into large boules, has excellent physical and electronic properties, and has the extraordinary attribute that a stable and electronically well-behaved oxide is easily formed on its surface. This oxide of silicon, SiO_2, is better known to most simply as glass or sand. A particular transistor

Figure 1. IC Feature Gate Size



device, the metal-oxide semiconductor field-effect transistor (MOSFET), and a particular circuit topology, namely the complementary metal oxide semiconductor (CMOS), have evolved to become the standard structures used in the scaling of the ULSI integrated circuit. No fundamentally new inventions were needed to undertake this scaling, which allowed industry to focus on the physics and fabrication science and technology of device scaling and not to be diverted by the need to invent a fundamentally new device.

Device scaling involves the scaling of the applied electric field, active channel lengths, junction dimensions and their electronic properties, insulator thickness, capacitance, and power dissipation—to name but a few of the geometric and material properties involved. The considerable innovation that has been required to accomplish device scaling should not be minimized. However, by maintaining focus on essentially one device type (although there were other device structures along the way, such as the bipolar transistor), one circuit topology (again, with several alternative topologies along the way, such as the n-channel MOS [NMOS], emitter-coupled logic [ECL], and integrated injection logic [I^2L]) and a limited number of materials, industry was able to focus its creativity and resources. This in no small measure contributed to the rapid rate of progress over the past 40 years. The ability to fabricate well-behaved transistor switches was due to the fact that the properties of these devices did not demonstrate significant quantum mechanical properties. This situation is now changing. Scaling based simply upon Moore's Law is projected to approach atomic dimensions by the year 2050. However, it is generally accepted that the current device physics paradigm will not permit CMOS switching transistors with well-behaved characteristics with feature sizes on the order of several atoms. In fact, the silicon IC electronics community recognizes that the present paradigm will encounter problems long before 2050.[3]

A brief review of several scaling arguments suggests some of the concerns. The detailed scaling of MOS devices is well understood but is technically complex and beyond the scope of this paper. However, many of the essential aspects can be understood

Timothy P. Coffey is a senior research scientist at the University of Maryland. He currently holds the Edison Chair for Technology at CTNSP. Gerald M. Borsuk is Superintendent of the Electronics, Science and Technology Division, at the U.S. Naval Research Laboratory.

from simple geometric arguments involving such elementary concepts as electrical resistance and capacitance. The resistance of a wire, for example, increases with its length and decreases with its cross-sectional area. The capacitance of a structure increases with its cross section and decreases with the separation between the plates of the capacitor. The devices (integrated circuits) we are interested in are predominantly resistive/capacitive systems connected by an array of switches (transistors). Such interconnected switching systems have time constants characterized by the product of the effective resistance and the effective capacitance (generally referred to as the RC time constant). These simple concepts are sufficient to gain a rudimentary understanding of what has transpired in the evolution of microelectronics over the years and some of the key issues facing this field at this time. Two of the principal features of solid-state electronics are the length scales involved and the voltages used. Let us suppose that we scale a selective length by a factor α such that a length L becomes L/α. Similarly, scale voltage by a factor β such that a voltage V becomes V/β. Applying these simple geometric arguments to the key properties of wire-connected CMOS switching circuits creates table 1.

Before using table 1 for predictive purposes, it is reasonable to ask if it accounts for past developments. For example, in 1970, the feature size was about $2x10^4$ nanometers (nm). In 2000, it was about $2x10^2$ nm, resulting in an $\alpha = 10^2$. Table 1 would therefore predict that the transistor density should have increased by 10^4 over this time. In 1970, Intel microprocessors had about 10^3 transistors per die; hence, the prediction of table 1 for the year 2000 would be about 10^7 transistors (10 million) per die. The actual number was about twice this value. This is in reasonable agreement, considering that chip areas more than doubled over this time. In 1970, ICs performed at clock rates (the rate at which a chip's logic elements are toggled) of about $2x10^3$ cycles per second; hence, the prediction for 2000 would be about $2x10^9$ cycles per second (~2 gigahertz [GHz]), which is approximately correct. Table 1 provides merely a back-of-the-envelope guide to scaling and should not be used for more than that. In particular, table 1 does not permit *ab initio* calculation of the parameters. The initial parameters must be provided from experimental observation.

Keeping in mind the above caveat, table 1 can be used to illustrate some of the challenges facing the semiconductor industry today. The transistor clock rate in the central processing unit core of a microprocessor is about three GHz. These microprocessors perform about eight floating point calculations per clock cycle. Let us conduct a simple gedanken experiment that ignores the potential sophistication of computational architecture improvements. Assume, for the purpose of discussion, that one could reduce the feature size to the ultimate limit of one atom (that is, ~0.27 nm for a silicon atom). Of course, the scaling shown in table 1 will break down well before this limit is reached. Therefore, the predictions do not represent reality. Nevertheless, they illustrate some of the problems that will be encountered as feature sizes approach the limit. Since feature sizes are about 130 nm in high-volume IC production, this size reduction would correspond to $\alpha = 481$. For this condition, table 1 predicts that microprocessor performance density would increase by a factor of ~100 million. The clock rate (assuming it scales as the transistor cutoff frequency) predicted by table 1 would be more than a terahertz.

Table 1: Geometric Scaling
(scale selected lengths by factor α; scale voltages by factor β)

Parameter	Scaling
Resistance at device level R_D	$R_{D0}\alpha$
Resistance of long wires R_W	$R_{W0}\alpha^2$
Capacitance at device level C_D	$C_{D0}\alpha^{-1}\ \kappa^*$
Capacitance of long wire C_W	C_{W0}
Charging time at local interconnect level τ_D	$R_{D0}C_{D0}\ \kappa$
Charging time at long wire τ_W	$R_{W0}C_{W0}\ \kappa\alpha^2$
Charging time through transistor τ_T	$\tau_{T0}\alpha^{-1}$
Transistor frequency f_T	α/τ_{T0}
Energy per unit area E_A	$E_{A0}\,\beta^{-2}\,\alpha$
Energy per unit volume E_V	$E_{V0}\,\alpha\,\beta^{-2}$
Power per unit area P_A	$P_{A0}\,\beta^{-2}\,\alpha^2$
Transistor density (# transistors per unit area) D_T	$D_{T0}\,\alpha^2$
Performance density (transitions/sec) $f_T D_T$	$f_{T0}\,D_{T0}\,\alpha^3$

* κ is scaled dielectric constant.

As a result, since instructions and data cannot move faster than the speed of light, one transistor at best could only interact with other logic elements that are within ~$2x10^{-2}$ centimeters (cm) of its location. For today's microprocessors operating at about 3 GHz, this maximum interaction length is on the order of 10 cm. This implies that, for today's chips, instructions and high-speed data can be moved across the entire chip each cycle. (In reality, other limitations, such as transistor current drive and interconnect line charging time, place more practical limits on transmission lengths of the highest speed signals for today's ULSI ICs.) Obviously, this architectural approach will break down long before reaching feature sizes of the order of one atom. Dealing with this issue will involve a substantial change in the prevailing paradigm (for example, moving to computing architectures that are highly local in character). While it may be possible to accommodate this need, the accommodation will be done by introducing greater complexity, thereby potentially jeopardizing the present scaling paradigm.

Another important parameter is power dissipation. The current GHz clock rate microprocessor IC dissipates about 40 watts/cm^2. Today's chips are operating at lower voltages, and microprocessor architectures in particular employ self-actuated power limiting features that turn off those circuits within the IC that are not involved in the function being performed. This strategy has dramatically reduced power dissipation and is being driven by low power portable devices, such as cellular phones and personal digital assistants (PDAs). To understand the issue on a more fundamental level table 1 shows that using constant voltage scaling ($\beta = 1$), the power dissipation of today's

microprocessor would increase to 18 megawatts (MW)/cm^2 for feature sizes of one atom. This exceeds the radiant energy at the surface of the sun and is obviously not an option. One solution that has been applied is to scale voltage down by some factor. For example, if constant electric field scaling (that is, $\beta = \alpha^{-1}$) is used, then table 1 predicts no increase in the power dissipation. This, however, would require that the devices be operated at the millivolt level and that other factors, such as static power dissipation due to tunneling currents, not be a significant factor. While operating voltages are being scaled down (from 5 volts to 3.3 volts to 1.8 volts to 1.0 volt, for example), there are practical and device physics limitations to such very low millivolt supply voltages. Reducing supply voltage for high performance transistors by a factor of 5 (that is, to about 0.5 volts predicted by the 2001 ITRS roadmap in 2013) reduces the power dissipation for our gedanken experiment to ~30 KW/cm^2—still a very large value.

Power management strategies using circuit architecture will not be sufficient as power dissipation continues to increase. Thus, it is quite likely that power management will become a limiting factor well before the ultimate feature size is reached. Figure 2 presents actual data[4] on the increase in energy dissipation over several recent generations of microprocessors and several additional projected future generations. The two limits of constant voltage scaling and constant electric field scaling are indicated on the figure starting at the 100-nm technology node. Clearly, constant voltage scaling is untenable, and constant electric field scaling is not achievable over the projected range of feature size. Reality will be some intermediate state between the two extremes (see the shaded area in figure 2). These data confirm that power management will become a major problem for the current paradigm in the near future. From only device physics consideration, the limit is likely to be set by thermal dissipation due to leakage current in devices that exhibit quantum tunneling in the gate insulator.

A similar analysis concerning clock rate as a function of technology node is plotted in figure 3. Historical data are taken from the performance of Intel microprocessors to the 180-nm technology node, while projections beyond this point are taken from the 2001 ITRS. The ITRS data start with the year 2001 and extend to 2016. During that period, the clock rate is given as 1.7 GHz in 2001 and projected to reach 28.75 GHz in the year 2016. A simple analysis of the data shows that the ITRS plot has a much lower slope (n = 1.52) than the slope of the historical microprocessor data (n = 2.3). It appears that the ITRS is projecting a significant slowdown in the rate of growth of the clock speed over historical data. Practical limitations, including power dissipation and circuit limitations, will likely further limit the existing paradigm as suggested in the figure.

To this point, we have confined our discussion to properties that can be attributed to bulk processes in semiconductors. However, more subtle processes are emerging. For example, while the current feature sizes shown in figure 1 are larger than those in which quantum (that is, not bulk) properties come into play, the insulator thickness required by the scaling is now about 20 atoms. This is well into the transition region where quantum tunneling currents are clearly observable. This is a qualitatively different situation from what has occurred in the past and jeopardizes scaling in the near term (over the next 5–10 years).

There may be partial solutions to this particular problem. It should be noted from table 1 that capacitance scales as the dielectric constant. This offers the possibility of obtaining the required gate oxide insulator capacitance per unit area by holding the oxide thickness constant (so as to avoid making the tunneling problem worse) and increasing the dielectric constant of the oxide. Therefore, the

Figure 2. Microprocessor Energy Density

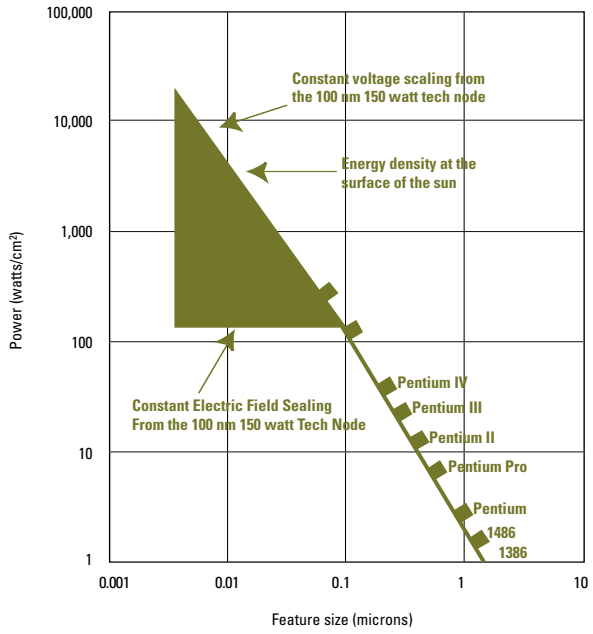
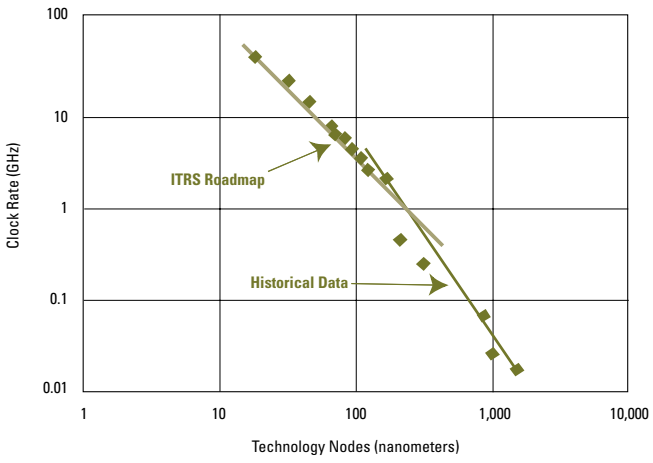


Figure 3. High Density Microprocessor Clock Rate as a Function of Technology Node
(after the ITRS 2001 Roadmap and historical data)



likely near-term solution to the insulator problem will be to find a new material system with a higher dielectric constant than silicon dioxide (SiO_2), so called high-κ materials, such that the electrical properties can be scaled without having to thin the insulators further.

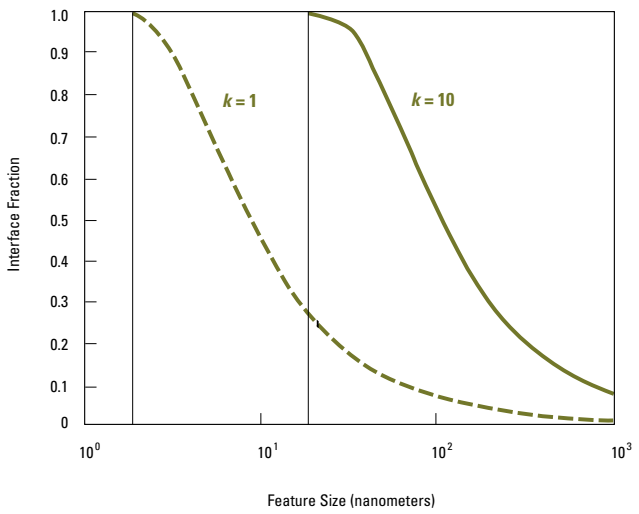
This is easier said than done, since the insulator material system must be compatible with the substrate upon which it is placed. In all likelihood, this problem will be resolved, since large resources will be applied to the solution. It does, however, make the point that the scaling is becoming much more complex than it has been. It should also be noted that this dielectric scaling strategy cannot be employed for more than one or two generations since it will ultimately destroy the geometry of the scaling (that is, the ratio of the channel length to the oxide thickness is typically at least 25 to 1 so that the gate can exercise its control function). Other innovations that will continue in the near term to keep the industry on the Moore curve are strained silicon for higher electron mobility in the active channel, high current/low voltage implantation for shallower junctions, and advances in copper/low-κ materials for interconnects, where κ here refers to dielectric contant.

However, one has to work harder and harder to maintain the scaling required by Moore's Law. Another way to get a sense that the situation is now qualitatively different is to estimate the fraction of atoms of a semiconductor feature that reside on the surface to those that reside in the volume (bulk) of the feature. An estimate of this fraction is shown in figure 4, where n is the feature dimension divided by the effective diameter of the atoms that make up the feature (that is, n is the number of atoms across the feature). The parameter κ is a measure of the number of atoms over which surface or interface effects manifest themselves. In the case $\kappa = 1$, these effects are manifest over only 1 atom thickness. In the case $\kappa = 10$, they are manifest over 10 atom thickness.

Today, feature sizes are approaching 100 nm, which corresponds to n = 370. It is clear from figure 4 that for values of n much larger than 370 (representing the previous history of solid-state electronics), the interface fraction is very small regardless of which value of κ one uses. For values of n smaller than 370 (the region into which scaling now enters), the interface fraction rises very rapidly until it becomes dominant. This is a very different situation from the past and will undoubtedly have profound effects on scaling to smaller feature sizes. When these effects occur depends upon the value of κ. If the effects are restricted to within 1 atom of the surface ($\kappa = 1$), they will not become appreciable until feature sizes reach about 30 nm. If the proper choice for κ is 10, then solid-state electronics is already into the domain where it is becoming dominated by interface processes rather than bulk processes. The appearance of quantum tunneling currents mentioned above suggests this case. Of course, this model is elementary, but it does suggest that obtaining another order of magnitude reduction in feature size will be considerably more challenging than it has been to date.

Because it is not possible to present a complete review of the technical literature, we will summarize a few concerns regarding continued scaling of semiconductor electronics. It has been pointed out that when the dimensions of devices approach the mean free path of the carriers (which is now occurring), the bulk transport models that have been used in the past fail. Quantization of the energy levels occurs when dimensions approach the deBroglie

Figure 4. Interface Fraction



wavelength of electrons. As dimensions approach the distance between the dopant atoms, small number effects will become important, because depleted layers must be reduced in proportion to the dimensions. Tunneling currents increase rapidly as layers are thinned. The scaling of the thickness of the gate insulation in field effect transistors in proportion to other dimensions of the transistor approaches limits set by tunneling and by the influence of the silicon–silicon dioxide interface on the insulating properties. The resulting power dissipation created by tunneling currents grows substantially and creates significant roadblocks to further higher-level integration. The breakdown voltage of insulating layers decreases. Soft errors, such as those created by radioactive impurities and cosmic rays, are becoming of increasing concern. The interconnect wiring between transistors is becoming a special concern.

Devices in computers are organized into logic circuits made of a few transistors that perform elementary logic functions. The circuits are interconnected to implement more complex functions. The number of connections is the same order as the number of components. The number of connections is now in the many millions and will grow exponentially in the near future. For example, the total length of interconnect wires in a modern microprocessor IC is several kilometers. Making wires narrower to reduce the space that they occupy on chips increases their resistance per unit length and leads to transmission delays limited by the time it takes to charge the capacitance of the wire given as 1/RC, where, again, R is the resistance and C is the capacitance of the wire. The wire length per transistor increases faster than the number of transistors. Deleterious cross-talk between more closely spaced interconnects also increases.

It is clear from this simple analysis that the wiring interconnects in ICs do not scale and thereby do create a significant barrier to further integration. All of the above contribute to increasing the complexity of feature size scaling and therefore increase the difficulty of maintaining the performance enhancements predicted by

Moore's Law. These roadblocks are well known to the microelectronics community. For example, the ITRS notes, "Processing dimensions are getting close to the size of photoresist molecules and other physical dimensions associated with exposure and development. Existing techniques for measuring sizes, positions, and defects are becoming difficult to use. In addition, displacement of the equipment's structural parts due to heat and vibration is no longer negligible." Clearly, there will be severe demands for metrology and stability if one is to reduce feature size further.

To accommodate Moore's Law in the near term, the issues discussed above and others must be managed such that the technology node (the ½ pitch size of the metal interconnect line width connecting transistors) approaches 65 nm by 2007. The physical gate length of the transistor for this technology node will be ~32 nm. Much beyond the 65-nm technology node, optical lithography will no longer be viable for additional feature size reduction. This will require that new lithographic technology is introduced if the current paradigm is to be pushed further. The most likely replacement is extreme ultraviolet (EUV) lithography, which operates at a source wave length of about 13 nm. EUV lithography is quite different from optical lithography. For example, it must use focusing mirrors rather than lenses. Also, it is absorbed by almost all materials, making masks very expensive and complex. This technology is likely to be much more expensive than existing optical lithography technologies that use laser sources at 193 nm and 157 nm and will continue the trend of ever more expansive manufacturing facilities.

The increasing costs for state-of-the-art semiconductor manufacturing facilities is also an important consideration that will shape the future direction of this technology. The discussion above has focused on physical limitations, but there is also an economic side, one manifestation of which is known as Moore's Second Law. This law is illustrated in figure 5, which plots the cost per factory as a function of time.[5] This figure illustrates the impact of increasing complexity on cost. Extrapolating the current trend to 2005, the cost per fabrication line will be about $10 billion, and by 2015 it will be at $200 billion.

Of course, industry will work to reduce these out year capital investments. Nevertheless, large capital investments clearly will be required to continue feature size scaling. The trend is toward fewer worldwide facilities using ever-larger wafer size (12-inch-diameter wafers are the state of the art today) to achieve economy of scale reductions. A corollary to current cost trends is that the ability to implement a diversity of chip designs into actual ICs will likely be constricted as the cost of a mask set for a given design becomes extremely high. The result for ICs with state-of-the-art features and density will likely be fewer different designs in extraordinarily high volume.

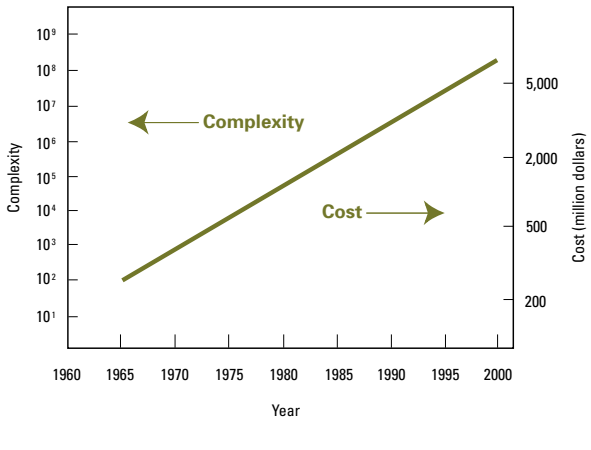
Is the market there to support such an investment, especially in light of the approaching physical limitations that will limit the long-term use of the new capital investment? Will any of the few mega-fabs of the future be in the continental United States? Will American companies own any of them? If the answer to either of these last two questions is no, then what are the implications for supply and assured chip functionality for DOD? Within about 15 years, as features approach the size of a few atoms, further miniaturization will not be possible. Will industry find it profitable to push the current paradigm to its limit? It is clear that to continue Moore's Law beyond 10 to 15 years will require the introduction of a new device and a new circuit topology paradigm. While it may be possible to do so, it must be realized that this will no longer be the development path that led to Moore's Law in the first place. Finding this new path, if it exists, becomes the key issue for the science and technology community that supports the IC manufacturing base.

It seems clear that solid-state microelectronics will enter a new regime over the next 7 to 10 years in which the current scaling paradigm will no longer hold

What Is Next?

It seems clear that solid-state microelectronics will enter a new regime over the next 7 to 10 years in which the current scaling paradigm will no longer hold. Interface and quantum mechanical processes rather than classical processes will dominate the emerging technological regime. Feature size scaling will become more difficult. Indeed, the scaling that has worked so effectively in the past will likely not convey to this new regime. It is reasonable to ask whether DOD should be concerned about this. One approach to coming to grips with this question is to look at where the current paradigm is likely to bottom out. It has been shown experimentally that the gate oxide thickness, which is the thinnest feature of the MOSFET, must be at least five atoms thick for silicon dioxide.[6] The experiments that determined this limit required extraordinary control in the creation of the oxide layer and used techniques that will not scale to mass-produced devices. Nevertheless, the experiments provide valuable guidance regarding how far the present paradigm can be pushed.

Figure 5. Moore's Second Law



The gate length in the MOSFET is typically 25 times the oxide thickness. This leads to a minimum gate length of ~125 atoms, or about 32 nm for a silicon dioxide insulator of about 1.2 nm for the 65-nm technology node expected in 2007. (Industry is prepared to make the investments needed to reach 65-nm feature size, and it should be reached by 2007.) It is logical to ask whether the 65-nm technology node would provide sufficient computer power to meet DOD needs. If this level would suffice, then DOD has no reason for concern. If additional computer power is required, what can be done to move beyond this point? The scaling in table 1 provides some guidance. If a conservative 3-year scaling rule methodology is used, then a feature size of 32 nm would correspond to about 775 million transistors per chip, of which ~276 million transistors would be high-performance devices in the core of the main power unit operating at a clock rate of ~7 GHz. Such an IC would have the ability to perform ~30 Giga FLOPS with a single microprocessor. It would also dissipate ~190 watts. This power dissipation, while stressing, is probably manageable, and the computer power is quite substantial.

A demanding military requirement is the DOD stated objective to maintain information superiority and total situational awareness

Is this computer power sufficient for expected DOD applications? To answer this requires some projection of future DOD requirements. A demanding military requirement in this regard is the DOD stated objective to maintain information superiority and total situational awareness. This objective will be accomplished partially by smart sensors and appropriate information networks. Since history suggests that we always need more compute power than we think, it is prudent to employ aggressive examples when making predictions. As an illustrative example, we examine a stressing but not unreasonable implementation of total situational awareness, namely target detection and identification using hyperspectral imagers flying on mini unmanned aerial vehicles (UAVs). The subject of using mini UAVs for such applications was recently addressed by one of the authors.[7]

The advantage of the mini UAV is that it can obtain high-resolution images and would be an organic asset of a local commander (for example, a battalion level commander). If the UAV flies at a 3-kilometer (km) altitude, seeks spatial resolution of about 6 inches, and flies a hyperspectral imager with a 20-degree field of view and 200 spectral bands, then each image looks at about 1 km^2 and generates about $2x10^8$ pixels per band. For convenience, assume a 16-bit dynamic range. The information content of a single image frame is then about 10^{11} bits. To achieve real-time battlefield assessment, it would be desirable to receive these images at video rates. This results in a bit stream of about $3x10^{12}$ bits per second. This sensor does not exist for reasons that will become clear shortly, as well as other issues such as the need to achieve the required signal-to-noise ratio. Nevertheless, it is instructive to look at where we stand regarding meeting the computational and data handling requirements necessary to operate on this raw data stream. The electronics to get the data off the focal plane and perform, for example, automatic target recognition (ATR) and data compression will require the image processor to process data at about the same rate that it is collected (that is, about 10^{12} operations per second[8]). This is about a hundred times the speed of the front side bus expected from a single microprocessor at the 65-nm technology node. One could segment the calculations so that they could be accomplished by multiple processors. This approach would require about a hundred microprocessors and would dissipate on the order of about 10 kilowatts. This is clearly beyond the capability of the mini UAVs, which are power constrained, and, therefore, the mission could not be done with this approach. This would also exceed the payload power available on larger UAVs, such as the Predator (~1.8 KW).

While this calculation is oversimplified, it does make the point that DOD information superiority and total situational awareness objective cannot be met if the effective computer power available to DOD distributed sensors asymptotes at the levels determined by the 32-nm MOSFET at the 65-nm technology node. In making this statement, we assume that the distributed sensors will need onboard processing power that is capable of dealing with the data collection rate. One could conceive of transmitting the uncompressed data to central facilities where detailed calculations would be done with supercomputers. This does not seem reasonable, however, considering the transmitter power that would be necessary, the data rates involved, and the operational environments in which the sensors would operate. We have also deliberately used the term *effective computer power* since there may be approaches other than brute force number crunching to deal with these high data rate sensor systems.

As mentioned earlier, the feature size of an MOS transistor can be reduced further by employing a gate insulator with a dielectric constant substantially higher than SiO_2 if a robust new material can be identified that has the necessary properties. For example, by doubling the dielectric constant of the gate insulator, a robust transistor with a gate length of 16 nm should be feasible. The resulting processor would still not meet the computing requirements outlined in the mini UAV example described above. Indeed, it appears that one would need a feature size of about 8 nm to do the job. At this point, the ratio of gate length to oxide thickness would be small enough that the device probably would not work. The power handling requirements would also be formidable for the current paradigm. Other DOD computing and data handling requirements can be identified that will require effective computing power that will probably not be achieved within the paradigm that has tracked Moore's Law over the past 40 years.

What Does the Future Hold?

Research is being conducted worldwide that offers fleeting glimpses of a new paradigm in solid-state electronics at the nanometer scale below 20 nm—the realm of *nanoelectronics*. This long-term research focuses on new materials and new electronic phenomena. For example, nanotubes of carbon and other materials have demonstrated amazing physical and electronic properties. The electronic properties—in some cases a metal and in others a semiconductor—of carbon nanotubes have been demonstrated to be a function of the

chirality of the nanotube and its diameter. Elemental transistor-like switches have been made from single carbon nanotubes. In a completely different research area, advances have been made in quantum dot structures. A zero-dimension quantum dot is a semiconductor structure that confines exactly one electron in all dimensions. These structures, made possible by advances in the deposition and precise control on an atomic level of thin layers of III–V semiconductors, have led to the discovery of so-called qubits—quantum dots that can have several different energy states simultaneously. The possibility of using such qubits in a quantum computer—a computer not based on Boolean logic but instead on a completely different mathematics with computational properties that are theoretically far beyond what exists today or will exist in traditional ULSI in the time horizon described herein—is intriguing.

Yet another area with potential for large impact in future nanoelectronics technologies is magnetic semiconductor materials and the possibility that such materials offer to create switches based upon controlling and sensing the quantum mechanical spin of single electrons. The concept of helical logic devices in which information-encoded single electrons are constrained to move along helical paths formed by a rotating electric field is yet another novel concept for advanced computing. Computing based upon biological implementations using deoxyribonucleic acid is also under intense study. In addition, there are potential breakthroughs in the area of processing algorithms. These examples, and those yet to be discovered, of high-risk long-term research are possible directions beyond the horizon of the present paradigm.

There is no promise, however, that any of these research areas will produce the answer. But if the history of 100 years of technological innovation in electronics is any guide, the prospects are in our favor that a way beyond the looming limitations of scaling and Moore's Law will be found. However, investment in this type of long-term high-risk research is unlikely to be made by the industrial private sector, which has more pressing near-term needs. Also, basic research is by its nature nonproprietary. The free exchange of ideas and results is a critical element of the scientific process. Research is a global enterprise. But there is a definite need to ensure that whatever new scientific breakthroughs occur, they are available first and foremost to the United States and DOD to maintain a technological advantage. The most assured way to have this occur is to nurture long-term research within the United States in the private sector, universities, and Government Laboratories.

A Course of Action

What is a prudent course of action for DOD? It would not seem productive for DOD to invest resources to help squeeze the most out of MOSFET CMOS scaling. Commercial industry has mastered this area and has the ability—and incentive—to apply large resources toward this problem. There is little that DOD can add except, perhaps, in niche research areas, such as advanced lithography. On the other hand, several fundamental device and circuit architecture issues have been outlined in the discussion above that any new paradigm must address satisfactorily. Among these are excessive power dissipation, very low-voltage operation, and interconnection and signaling limitations as switching rates increase. These and other related technological issues require new innovations in material science, fabrication, and architectural approaches for their resolution.

A sustained investment in these areas is where DOD can once again make a major impact on future electronics for sensing, computation, and information technologies. The highest leverage in development programs occurs in the early stages of research and development, where investment costs are low and the opportunity for impact is high. This is where DOD science and technology can make significant contributions. Undoubtedly there are many other approaches that have not yet been thought of or surfaced. Clearly there is great opportunity here for DOD science and technology to make major contributions. Also, it is important to keep in mind the discovery that led to the invention of the transistor. Breakthroughs in one area often result from serendipitous discovery in what appear to be unrelated areas. Positioning DOD so that it can maintain the needed broad visibility in the technical community and to be wise enough to recognize important developments/discoveries will be key to success.

Notes

[1] G.E. Moore, " Cramming more components onto integrated circuits," *Electronics* 38, no. 8 (April 19, 1965).

[2] R.H. Dennard, F.H. Gaensslen, H.N. Yu, V.L. Ridout, E. Bassous, and A.R. LeBlanc, "Design of ion-implanted MOSFETs with very small dimensions," *IEEE Journal of Solid-State Circuits* 9 (October 1974), 256.

[3] Community awareness of this problem is indicated by publications such as J.D. Meindl, ed.," Special Issue on Limits of Semiconductor Technology," *Proceedings of the IEEE* 89, no. 3 (March 2001), and P. M. Solomon, ed., "Scaling CMOS to the limit," *IBM Journal of Research and Development* 46, no. 2/3 (March/May 2002), 117–360.

[4] S. Borkar, "Design challenges of technology scaling," *IEEE Micro* 19 (July/August 1999), 23–29.

[5] E.S. Meieran, "21st Century Semiconductor Manufacturing Capabilities," *Intel Technology Journal* (4th quarter 1998).

[6] D. Muller et al., *Nature* (June 24, 1999).

[7] Timothy Coffey and John Montgomery, "The Emergence of Mini UAVs for Military Applications," *Defense Horizons* 22 (Washington, DC: National Defense University Press, December 2002).

[8] This number depends on the algorithms used.

Defense Horizons *is published by the Center for Technology and National Security Policy through the Publication Directorate of the Institute for National Strategic Studies, National Defense University.* **Defense Horizons** *and other National Defense University publications are available online at http://www.ndu.edu/inss/press/nduphp.html.*

Center for Technology and National Security Policy

Hans Binnendijk
Director